1939

Use of functionals in obtaining approximate solutions of linear operational equations

George Lloyd Gross
Iowa State College

USE OF FUNCTIONALS IN OBTAINING APPROXIMATE SOLUTIONS OF LINEAR OPERATIONAL EQUATIONS

by

George Lloyd Gross

A Thesis Submitted to the Graduate Faculty
for the Degree of

DOCTOR OF PHILOSOPHY

Major Subject Applied Mathematics

Approved:

Signature was redacted for privacy.

In charge of Major work

Signature was redacted for privacy.

Head of Major Department

Signature was redacted for privacy.

Dean of Graduate College

Iowa State College

1939

TABLE OF CONTENTS

I. INTRODUCTION

In many situations one is able to formulate a problem mathematically, but even if methods of exact solution exist, the answer obtained may be too cumbersome for interpretation or the operations used in solving may become too involved for actual accomplishment. Similar troubles are sometimes encountered in the use of known methods of approximate solution. However, methods of approximation inject new versatility into the solution and, after all, the analytical expression of the problem is no more than an approximation of the real situation.

Many problems are special cases of a more general linear operational equation. Several of the methods used in solving these special problems with various degrees of approximation are examples of the more extensive method used in this paper in obtaining an approximate solution of the general linear operational equation. It is necessary in inexact methods not only to arrive at a result by practical means but also to be able to study the resulting degree of approximation.

This paper explains the use of functionals in obtaining an approximation of the solution, explores ways of estimating the error involved, relates the method to existing methods, and indicates a number of ways in which extensions can be made.

II. REVIEW OF LITERATURE

A linear operational equation is a generalization of the form of a large class of problems, some of which have been solved with various degrees of approximation by different investigators.

When the boundary conditions of a problem are not very complicated, it may happen that functions satisfying all of them simultaneously can be found. The Ritz (1) method is then frequently very practical in approximating the solution. The methods of Boussinesq (2) (least squares) and of Krawtchouk (3) are also applicable to the same type of problem.

The wave-mechanical perturbation theory (4) developed by E. Schrödinger (5) and a generalized perturbation theory (4) developed by P. S. Epstein (6) have both been successfully applied to many problems of quantum mechanics. Perturbation theory is applicable when another problem with the same boundary conditions can be solved.

The Trefftz (7) method of approximating the solution of a problem can often be used when exact solutions of the differential equation are known. One is usually forced to weaken the boundary conditions in applying the method.

The method of functionals generalizes these methods and permits progress in the solution of a more general problem.

III. INVESTIGATION

A. The Problem $LU=P$.

1. Equation with Linear Operators.

In the problem $LU=P$ the $m\times 1$ matrix P with functions as elements is given, the unknown $n\times 1$ matrix U with functions as elements is to be determined, and the $m\times n$ matrix L with linear operators as elements is given. The elements of the matrix L commute with constants and transform to zero expressions vanishing identically.

If the relation $LU=P$ is expanded in terms of the elements of the matrices L, U, and P, there results the system of equations,

$$\begin{aligned}
&L_{11}U_1 + L_{12}U_2 + \cdots + L_{1n}U_n = P_1 \quad ,\\
&L_{21}U_1 + L_{22}U_2 + \cdots + L_{2n}U_n = P_2 \quad ,\\
&\cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad ,\\
&L_{m1}U_1 + L_{m2}U_2 + \cdots + L_{mn}U_n = P_m \quad .
\end{aligned}$$

In terms of the usual summation convention, the same problem can be written in the form,

$$L_{ij}U_j = P_i \quad ,$$
$$(i=1,2,\ldots,m) \quad ,$$
$$(j=1,2,\ldots,n) \quad .$$

If one desires to emphasize the property of matrix multiplication, one can write the same thing in the following way,

$$\begin{bmatrix} L_{11} & L_{12} & \cdots & L_{1n} \\ L_{21} & L_{22} & \cdots & L_{2n} \\ \cdot & \cdot & \cdot & \cdot \\ L_{m1} & L_{m2} & \cdots & L_{mn} \end{bmatrix} \begin{bmatrix} U_1 \\ U_2 \\ \cdot \\ \cdot \\ \cdot \\ U_n \end{bmatrix} = \begin{bmatrix} P_1 \\ P_2 \\ \vdots \\ P_m \end{bmatrix} .$$

The brief matrix form of expression $LU=P$ is generally used in the following material.

2. <u>Examples of Problem</u>.

a. <u>Non-homogeneous</u>. If the matrix equation $P=0$ does not hold, the problem is non-homogeneous.

The system of linear differential equations,

$$\left(\frac{d^2}{dt^2} - 4\frac{d}{dt} + 4\right)x - y = 0 \quad ,$$
$$25x - \left(\frac{d^2}{dt^2} + 4\frac{d}{dt} + 4\right)y = -16e^t \quad ,$$

illustrates the non-homogeneous problem. Another example is the expansion of functions such as of $\sin x$,

$$U = \sin x .$$

The linear integral equation

$$\left\{1 - \int_a^x K(x,s)\left[1(\,)\right]_{x=s} ds\right\} \varphi(x) = f(x)$$

is evidently a problem of the same kind.

Indefinite integration often can be put in the form of the non-homogeneous problem. One can, for example, write

$$\frac{d}{dx} U = \frac{\sin x}{x}$$

$$[1()]_{x=0} U \equiv [U]_{x=0} = 0$$

in place of

$$U = \int_0^x \frac{\sin x}{x}\, dx .$$

Linear boundary value problems, such as the problem of determining $U(x,y)$ to satisfy simultaneously the differential equation

$$\left(\frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2}\right) U = 4\pi\rho(x,y)$$

and the boundary condition

$$[1()]_{g(x,y)=0} U = 0 ,$$

are further examples.

Many other examples can be given. Certain linear difference equations, systems of linear integral equations, linear integro-differential equations, and systems of linear operational equations are of the form $LU = P$.

b. <u>Homogeneous</u>. If the relation $P = 0$ holds, the problem is said to be homogeneous. Examples corresponding to those given above can be formulated. The system of linear differential equations

$$\begin{aligned} \frac{d}{dt} x - 2y \qquad\qquad &= 0 , \\ \frac{d}{dt} y - 2z &= 0 , \end{aligned}$$

$$-2x \qquad + \frac{d}{dt}z = 0 \quad ,$$

and the linear integral equation

$$\left\{\int_a^x K(x,s)\,[1()]_{x=s}\,ds\right\}\varphi(x) = 0$$

immediately bring many to mind.

c. <u>Characteristic Value</u>. Examples of characteristic value problems are the linear integral equation

$$\left\{1-\lambda\int_a^b K(x,s)[1()]_{x=s}\,ds\right\}\varphi(x) = f(x) \quad ,$$

the Schrödinger wave equation

$$(H-\lambda)\Psi = 0$$

of quantum mechanics, and the vibrating string problem

$$\left(\frac{\partial^2}{\partial x^2}+\lambda^2\right)U = 0$$

$$[1()]_{x=0}\,U = 0$$

$$[1()]_{x=1}\,U = 0 \;.$$

3. <u>Generalization to a Problem $E(U;K)=0$</u>.

While the present paper is mainly concerned with the treatment of the linear problem $LU=P$, this is a special case of the problem

$$E(U;K)=0 \;,$$

the letters K, U, and E respectively symbolizing known elements, unknown elements defined by the relation $E=0$, and a known expression involving the elements K and U. A treatment of this more general problem is now sketched after the manner in which the linear problem is treated.

In order to obtain an approximate solution one substitutes in the relation $E(U;K)=0$ the expression $e(u;k)$ for the unknown U, the letters k, u, and e respectively symbolizing known elements, unknown elements to be determined, and a known expression involving the elements k and u. It is not, in general, to be expected that the expression e can satisfy the relation, and one writes

$$0 \cong E(e;K) = E(e(u;k);K) \equiv \mathcal{E}(u;k;K).$$

One performs the same operation F on the expression $\mathcal{E}$ and on zero, the operation being of such a character that, upon equating the two results of the operation, a determination, say d, for the unknown u is obtained. In other words, from the relation $F(\varepsilon)=F(0)$ one gets a result d for the unknown u.

One has the known expression $e(d;k$ for the unknown U. The expression e in general does not satisfy the relation $E=0$ and one then studies ways of measuring or judging the error introduced.

The linear problem $LU=P$ is not only an important but also a tractable case of the problem $E(U;K)=0$. In the latter case one can not always hit upon an operation F that permits a determination of the unknown u. In the subsequent treatment of the linear problem in a like manner this difficulty does not arise.

B. Determination of an Approximation.

The approximate solution of the problem $LU=P$ by the method of functionals is given first without the details. One substitutes for the unknown a product VA of a known matrix V by an unknown constant matrix A, and obtains $L(VA)\cong P$. It is the sense of the present method of approximation that, while the foregoing equation is not exactly true, the constant matrix is determined by an equation of the form

$$F(L(VA))=FP$$

in which F is a functional.

The matrix U is replaced by the matrix VA and, in full, one has

$$\begin{bmatrix} U_1 \\ U_2 \\ \vdots \\ U_n \end{bmatrix} \cong \begin{bmatrix} V_1 & 0 & \cdots & 0 \\ 0 & V_2 & \cdots & 0 \\ \cdot & \cdot & \cdot & \cdot \\ 0 & 0 & \cdots & V_n \end{bmatrix} \begin{bmatrix} A_1 \\ A_2 \\ \vdots \\ A_n \end{bmatrix}$$

in place of $U\cong VA$. The unknown $n\times 1$ matrix A, with constants as elements, is to be determined and the $n\times n$ matrix V, with selected functions of expansion as elements, is known.

Each element of V and of A is a matrix. Since the elements of one matrix, say V_1, are not necessarily the same as those of another, say V_2, let k_j be the number of elements of V_j and, in full, one has

$$(j = 1, 2, \ldots, n)$$

$$V_j = [V_{j1} \; V_{j2} \; \cdots \; V_{jk_j}]$$

$$A_j = \begin{bmatrix} A_{j1} \\ A_{j2} \\ \vdots \\ A_{jk_j} \end{bmatrix} .$$

Let s be the total number, the sum $k_1 + k_2 + \cdots + k_n$, of elements of V or of A. Since the elements of V_j may not be linearly independent, let l_j be the number of elements that are. Naturally the inequality $l_j > k_j$ is impossible. Let l be the number of elements of V that are linearly independent. Evidently one has $l \leqq l_1 + l_2 + \cdots + l_n \leqq s$.

If one replaces the unknown U by the expansion VA in the equation $LU = P$, one cannot expect, in general, the expression $L(VA)$ to equal the other member P for any value of A. One gets $L(VA) \cong P$, and this is an attempt to represent the matrix P as the product of the known matrix LV by a constant matrix A. One can verify by expansion of the various products involved that, in spite of the operational property of L, the usual associative property of matrices holds for the above products. In particular, one can show

$$L(VA) = LVA = (LV)A \quad .$$

Let r be the number of the s columns of the matrix that are linearly independent. The number r cannot be greater than the sum $l_1 + l_2 + \cdots + l_n$ because of the linearity of the operator L.

An $m \times m$ matrix F with linear functionals as elements is selected. The elements of the matrix F also commute with constants and transform to zero expressions vanishing identically. For simplicity it is often convenient to have all non-diagonal terms vanish so that the matrix F has the form,

$$\begin{bmatrix} F_1 & 0 & \cdots & 0 \\ 0 & F_2 & \cdots & 0 \\ \cdot & \cdot & \cdot & \cdot \\ 0 & 0 & \cdots & F_m \end{bmatrix} .$$

The selection is made such that the products $F(L(V))$ and $F(P)$ completely expanded are $r \times s$ and $r \times 1$ matrices with known constants as elements, and the rank of the matrix $F(L(V))$ is r, the number of linearly independent columns of the matrix LV.

Here each element F_i of F is a $g_i \times 1$ matrix. This symbolizes the use g_i times of the i-th relation of $L(VA) \cong P$. The numbers g_i are taken such that the sum $g_1 + g_2 + \cdots + g_m$ is r.

The relation $F(L(VA)) \cong FP$ can be replaced by the matrix equation $F(L(VA)) = FP$. The r equations in the s unknowns are not inconsistent. Because there are only r equations, the rank of the augmented matrix cannot exceed r. The rank of the augmented matrix must be at least r, because the rank of the coefficient matrix is r. It follows that the rank of the augmented matrix too is r.

Let the equation

$$(G(M(W)))B = GQ$$

represent the result after a transposition of the terms involving $s-r$ of the unknowns of the member $(F(L(V)))A$ to the other member FP. The rank of the matrix GMW can and must be r so that $(GMW)^{-1}$ exists. The matrix B lacks $s-r$ of the elements of A, but all r of the elements of the former are elements of the latter. Each element of G, M, and W is one of the elements of F, L, and V respectively. Each element of GQ is, except for the addition of certain transposed terms, an element of FP.

The notation change from F, L, V, A, P to G, M, W, B, Q is to take care not only of the transposition of certain terms but also to result in matrices without matrices as elements. The matrices, fully expanded for inspection, appear in the following equation

$$\begin{bmatrix} G_1 & 0 & \cdots & 0 \\ 0 & G_2 & \cdots & 0 \\ \cdot & \cdot & \cdot & \cdot \\ 0 & 0 & \cdots & G_r \end{bmatrix} \begin{bmatrix} M_{11} & M_{12} & \cdots & M_{1r} \\ M_{21} & M_{22} & \cdots & M_{2r} \\ \cdot & \cdot & \cdot & \cdot \\ M_{r1} & M_{r2} & \cdots & M_{rr} \end{bmatrix} \begin{bmatrix} W_1 & 0 & \cdots & 0 \\ 0 & W_2 & \cdots & 0 \\ \cdot & \cdot & \cdot & \cdot \\ 0 & 0 & \cdots & W_r \end{bmatrix} \begin{bmatrix} B_1 \\ B_2 \\ \vdots \\ B_r \end{bmatrix} = \begin{bmatrix} G_1 & 0 & \cdots & 0 \\ 0 & G_2 & \cdots & 0 \\ \cdot & \cdot & \cdot & \cdot \\ 0 & 0 & \cdots & G_r \end{bmatrix} \begin{bmatrix} Q_1 \\ Q_2 \\ \vdots \\ Q_r \end{bmatrix} .$$

There are now r equations in r unknowns and the coefficient determinant is not zero. The inverse $(G(M(W)))^{-1}$ exists and solution of the equations gives

$$B = (G(M(W)))^{-1}(GQ).$$

The elements of B are determined uniquely in terms of the transposed $s-r$ elements of A. The solution for B is complete. The corresponding r elements of A are uniquely determined in terms of the $s-r$ arbitrary elements of A. The solution for A is complete.

One can show that the preceding products are associative. If the parentheses are removed, the consecutive steps in the approximation are the following:

$$LU = P \quad ,$$
$$U \cong VA \quad ,$$
$$LVA \cong P \quad ,$$
$$FLVA = FP \quad ,$$
$$GMWB = GQ \quad ,$$
$$B = (GMW)^{-1} GQ \quad .$$

C. Estimation of the Error.

The equations

$$FLVA = FP$$

and the equations

$$GMWB = GQ$$

are satisfied exactly by the approximation of the preceding section. One can sometimes infer from this fact and from the character of the functional F how nearly the relations $LVA \cong P$ and $MWB \cong Q$ approximate equations. The same fact gives zero value for any linear combination of the elements of $F(LVA-P)$ or of $G(MWB-Q)$ by the linear property of F. An inference as to the value of $LVA-P$ or of $MWB-Q$ can sometimes be made from the character of F.

If one defines a remainder J by the equation

$$J = P - LVA \quad ,$$

one obtains the equality

$$FJ = 0 \quad .$$

In other words the remainder is such that, operated upon by the functional selected, the product vanishes. Knowledge of the character of the functional matrix can conceivably indicate the properties of the remainder.

The measure of error in the relation $Q \cong MWB$ is

$$Q-MWB=\frac{\begin{vmatrix} Q & MW_1 & MW_2 & \cdots & MW_r \\ G_1Q_1 & G_1M_{11}W_1 & G_1M_{12}W_2 & \cdots & G_1M_{1r}W_r \\ G_2Q_2 & G_2M_{21}W_1 & G_2M_{22}W_2 & \cdots & G_2M_{2r}W_r \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ G_rQ_r & G_rM_{r1}W_1 & G_rM_{r2}W_2 & \cdots & G_rM_{rr}W_r \end{vmatrix}}{\begin{vmatrix} G_1M_{11}W_1 & G_1M_{12}W_2 & \cdots & G_1M_{1r}W_r \\ G_2M_{21}W_1 & G_2M_{22}W_2 & \cdots & G_2M_{2r}W_r \\ \cdot & \cdot & \cdot & \cdot \\ G_rM_{r1}W_1 & G_rM_{r2}W_2 & \cdots & G_rM_{rr}W_r \end{vmatrix}}.$$

Every element of this fractional expression is known as soon as the matrices L, V, P, and F and the change from F, L, V, P to G, M, W, Q are selected. Although this fractional expression is really the value of $Q-MWB$ after the substitution of $(GMW)^{-1}GQ$ for B, the actual solution for B and A need not be carried out before the expression is studied. Knowledge of the change from F, L, V, P to G, M, W, Q enables one to write a similar fractional expression for $P-LVA$.

Assume that there exists a constant matrix a such that $U=Va$. One obtains the equations

$$LVa = P,$$
$$FLVa = FP,$$
$$GMWb = Gq,$$

$$b = (GMW)^{-1} Gq \, ,$$

by the same procedure used in finding $B = (GMW)^{-1} GQ$. One has

$$B - b = (GMW)^{-1} G(Q - q) \quad ,$$

the quantities Q and q differing at most in the elements of A chosen arbitrarily and the values given them. If these are the same,

$$B = b$$

and

$$A = a \qquad .$$

Assume that there exists a constant matrix a such that $U = Va + \varepsilon$ and something is known about the matrix ε. One obtains the equations

$$LVa + L\varepsilon = P \; ,$$
$$FLVa + FL\varepsilon = FP \, ,$$
$$GMWb + GM\eta = Gq \; ,$$
$$b = (GMW)^{-1} Gq - (GMW)^{-1} GM\eta \quad ,$$

in the usual manner. One has

$$B - b = (GMW)^{-1} G(Q - q) + (GMW)^{-1} GM\eta \quad .$$

For $Q = q$ one has the additional relations

$$B - b = (GMW)^{-1} GM\eta \equiv \mu \quad ,$$
$$WB - Wb = W\mu \quad ,$$
$$Q - MWB = M\eta - MW\mu \quad ,$$

and the quantities

$$A - a \; ,$$

$$VA - Va \ ,$$

are determined. If the matrix ε is zero, one has the results of the preceding paragraph.

Assume that there exists a constant matrix a such that $LVa + \bar{\varepsilon} = P$ and something is known about the matrix $\bar{\varepsilon}$. The equations

$$LVa + \bar{\varepsilon} = P \ ,$$
$$FLVa + F\bar{\varepsilon} = FP \ ,$$
$$GMWb + G\bar{\eta} = Gq \ ,$$
$$b = (GMW)^{-1}Gq - (GMW)^{-1}G\bar{\eta} \ ,$$

are obtained in the usual way. One has

$$B - b = (GMW)^{-1}G(Q-q) + (GMW)^{-1}G\bar{\eta} \ .$$

For $Q = q$ one has the additional relations

$$B - b = (GMW)^{-1}G\bar{\eta} \equiv \bar{\mu} \ ,$$
$$WB - Wb = W\bar{\mu} \ ,$$
$$Q - MWB = \bar{\eta} - MW\bar{\mu} \ ,$$

and the quantities

$$A - a \ ,$$
$$VA - Va \ ,$$

are determined.

D. Characteristic Values in the Problem.

Let the homogeneous problem $LU = P \equiv 0$ with a characteristic number λ in the linear operator L be considered. Assume V is such that, if and only if λ has certain values $\lambda^{(i)}$, is there an A, say $A^{(i)}$, such that one has

$$L^{(i)} V A^{(i)} = 0 \quad ,$$

the symbol $L^{(i)}$ signifying the operator L is evaluated for $\lambda = \lambda^{(i)}$. If λ has such a value, the elements of LV are linearly dependent and, because linear dependence is preserved by the linearity of F, the dependence is obtained in the solution by the method of functionals. If λ has some other value, the elements of LV are linearly independent and, because a functional is selected so as to preserve linear independence, one gets a trivial solution $A = 0$. The correct values of λ are obtained and, consequently, the correct values of A are arrived at.

When V is not sufficient to permit an exact solution of the problem of the type VA, one can still get results for the characteristic values. Of course, the degree of approximation can often be judged by substituting the result obtained in the equation $LU = P$, but this is not very satisfactory.

E. Relation to Other Methods.

Many methods of approximate solution have been given for the problem $LU = P$ of the special form

$$\begin{bmatrix} L_{11} \\ L_{21} \\ L_{31} \\ \cdots \\ L_{m1} \end{bmatrix} \begin{bmatrix} U_1 \end{bmatrix} = \begin{bmatrix} P_1 \\ 0 \\ 0 \\ \cdots \\ 0 \end{bmatrix} .$$

Let it be assumed that the solution can be approximated by the expression $V_1 A_1$. The matrix V_1 is selected so that each of its elements V_{11}, V_{12}, ..., V_{1k_1} satisfies all of the conditions of the problem except the relation $L_{11} U_1 = P_1$, which is the only one not satisfied by the expression $V_1 A_1$, with A_1 arbitrary. It is unfortunate that it is often difficult to transform the general problem to one with not more than one non-vanishing element in the matrix P. Except for rather simple problems it is impossible to determine elements for V_1 with the necessary characteristics. However, if such elements can be found, one assumes

$$U_1 \cong V_1 A_1$$

and

$$L_{11} V_1 A_1 \cong P_1 .$$

The above procedure with the functional F defined in the special way

$$F_1(\quad) = \int V_1'(\quad)\, d\tau_1$$

is equivalent to the Ritz (1) method. The prime over a matrix indicates its transpose is to be taken, and the region

τ_1 is the one throughout which the relation $L_{11}U_1 = P_1$ is to hold. Ritz treated problems with homogeneous boundary conditions and approximately satisfied the differential equation by obtaining through the variational method the same equations for his determination of the coefficients.

If the functional F is chosen in the form

$$F_1(\ \ ) = \int [L_{11}V_1]'(\ \ )\,d\tau_1 \quad ,$$

one gets the same equations for the determination of the coefficients as in the method of Boussinesq (2) (method of least squares).

The operator L_{11} can be written in infinitely many ways (3) in the form $\bar{L}_{11} + \lambda \hat{L}_{11}$ with λ a constant and $\bar{L}_{11}$ and $\hat{L}_{11}$ known operators. If the functional F is taken in the following way

$$F_1(\ \ ) = \int [\bar{L}_{11}V_1]'(\ \ )\,d\tau_1 \quad ,$$

one arrives at the equations of the method of Krawtchouk (3) for his determination of the coefficients of expansion. When the constant λ is zero, the method of Krawtchouk agrees with that of least squares.

In order to secure agreement with the generalized perturbation theory (4) of Epstein (6) and the wave-mechanical perturbation theory (4) of Schrödinger (5), one takes for F the functional

$$F_1(\ ) = \int V_1'(\ )\, J\, d\tau_1 \quad .$$

The function J is known in the region τ_1. Non-degenerate perturbation theory of the first and higher orders and degenerate perturbation theory are some of the special cases of this.

In the Trefftz (7) method the differential equation is satisfied by each term of an approximate solution of a boundary condition. The functional selected is

$$F_1(\ ) = \int Y(\ )\, d\tau_1 \quad ,$$

and the equations of Trefftz determining the coefficients are obtained. The symbol Y usually represents a normal derivative of V_1'.

If one specializes the problem $LU = P$ to the form

$$[U_1] = [P_1] \quad ,$$

there is obtained the problem of expansions of a given function P_1 in terms of known functions V_1.

If one expands the function

$$P_1(x_1, x_2, \ldots, x_n)$$

in a Taylor's series

$$\sum_{m_1} \cdots \sum_{m_n} (x_1 - h_1)^{m_1} \cdots (x_n - h_n)^{m_n} A_{m_1 \ldots m_n} \quad ,$$

the determination of the elements of A can be obtained by the method of functionals with the selection of F in the fashion

$$F_{i_1 \cdots i_n}(\quad) = \left[\frac{\partial^{i_1}}{\partial x_1^{i_1}} \cdots \frac{\partial^{i_n}}{\partial x_n^{i_n}}(\quad)\right]_{\substack{x_1=h_1 \\ \cdots \\ x_n=h_n}} .$$

If one expands the function P_1 in terms of orthogonal functions V_1, the functional chosen is

$$F_1(\quad) = \int V_1'(\quad) J \, d\tau_1 \qquad .$$

For convenience, J and V_1 are usually chosen with the property

$$\int V_1' V_1 J \, d\tau_1 = I \quad ,$$

the matrix I being the unit matrix.

The Ritz and allied methods are fundamentally expansions of the known function P_1 in terms of known functions $L_{11}V_1$.

An expression for U_1 in the form of a product of a known factor by an expansion is sometimes assumed. This product can be assumed of the form V_1A_1 as a simple example shows. Imagine the problem

$$L_{11}U_1 = P_1 \quad ,$$

$$[1(\,)]_{x=a} U_1 = 0 \quad ,$$

$$[1(\,)]_{x=-a} U_1 = 0 \quad .$$

If there is assumed the approximation

$$U_1 \cong (x-a)^2 (V_{11}A_{11} + V_{12}A_{12} + \cdots + V_{1k_1}A_{1k_1}) \quad ,$$

in order to satisfy the last two conditions of the problem, it is obviously the same as assuming the representation

$$U_1 \cong (x-a)^2 V_{11} A_{11} + (x-a)^2 V_{12} A_{12} + \cdots + (x-a)^2 V_{1k_1} A_{1k_1},$$

which is of the form $U_1 \cong V_1 A_1$ with a new definition of V_1 .

F. Extensions of the Method.

1. Biorthogonalization.

Whenever one gets a determination $B = (GMW)^{-1} GQ$ for the sum WB, biorthogonalization is possible. If one gets an answer using the functional G, there exists a functional $\mathfrak{G}$ such that the unknown matrix B is the product $\mathfrak{G}Q$ and the elements of $\mathfrak{G}$ are linear combinations of those of G. The expression $\mathfrak{G}$ has the form $(GMW)^{-1} G$, because one can obtain the result in the following way,

$$\begin{aligned} \mathfrak{G} MWB &= \mathfrak{G} Q \\ (GMW)^{-1} GMWB &= \\ IB &= \\ B &= \mathfrak{G} Q \quad . \end{aligned}$$

The idea of the functional $\mathfrak{G}$ is of no advantage in practical work. If one knows the matrix $\mathfrak{G}$, its application is still rather involved in the usual case in which $\mathfrak{G}$ is no longer diagonal but of the form

$$\mathcal{G} = \begin{bmatrix} \mathcal{G}_{11} & \mathcal{G}_{12} & \cdots & \mathcal{G}_{1r} \\ \mathcal{G}_{21} & \mathcal{G}_{22} & \cdots & \mathcal{G}_{2r} \\ \cdot & \cdot & \cdot & \cdot \\ \mathcal{G}_{r1} & \mathcal{G}_{r2} & \cdots & \mathcal{G}_{rr} \end{bmatrix} .$$

If one does not know the matrix $\mathcal{G}$, the determination of the elements of $\mathcal{G}$ or $(GMW)^{-1}G$ is comparable with the determination of B when G is used.

2. Analogue of Orthogonalization of Functions.

When a product WB is obtained by the method of functionals in the manner described, the addition of a linearly independent column to the matrix MW means not only determining the corresponding additional element of B but also a necessary alteration of the value of each element of B previously found.

In place of the matrix W take the product WC, say X. The constant matrix C is determined by imposing the condition

$$GMX = \bar{\delta} \quad ,$$

the value of $\bar{\delta}$ being taken in a particular way

$$\bar{\delta} = \begin{bmatrix} 1 & 0 & 0 & \cdots & 0 \\ G_2M_{21}X_1 & 1 & 0 & \cdots & 0 \\ G_3M_{31}X_1 & G_3M_{32}X_2 & 1 & \cdots & 0 \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ G_rM_{r1}X_1 & G_rM_{r2}X_2 & G_rM_{r3}X_3 & \cdots & 1 \end{bmatrix} .$$

One has the equations

$$\begin{aligned} \bar{\delta} &= GMX \quad , \\ &= GMWC \quad , \end{aligned}$$

and the second gives $C=(GMW)^{-1}\bar{\delta}$. The matrix C exists, because the inverse $(GMW)^{-1}$ is assumed to exist and the other factor $\bar{\delta}$ gives no trouble.

If the functional G is applied to the representation $MX\bar{B}$ instead of to MWB, the equations

$$\begin{aligned} GMX\bar{B} &= GQ \quad , \\ \bar{\delta}\bar{B} &= GQ \quad , \end{aligned}$$

result. The matrix $\bar{\delta}$ in the second has at least some of the advantages the matrix I would have. The solution $\bar{B}=\bar{\delta}^{-1}GQ$ exists, because the inverse $\bar{\delta}^{-1}$ is seen by inspection to exist. It is, however, unfortunate that $\bar{\delta}$ is not symmetric in all cases.

The preceding generalization has advantages in computation. The addition of a linearly independent column to the product MW means the calculation of only the corresponding additional element of B. An error in calculating one element of B in this manner would not change the values of the previously determined elements of B. However, the same answer is obtained both ways, and a verification of this fact is given in the series of equations,

$$
\begin{aligned}
MX\bar{B} &= MX\bar{\delta}^{-1}GQ \,, \\
&= MWC\bar{\delta}^{-1}GQ \,, \\
&= MW(GMW)^{-1}\bar{\delta}\bar{\delta}^{-1}GQ \,, \\
&= MW(GMW)^{-1}GQ \,, \\
&= MWB \,.
\end{aligned}
$$

The analogue parallels the orthogonalization (8) of functions very closely. In fact, the orthogonalization process is obtained by a certain choice of the matrices in the analogue.

3. Superposition of Approximations.

If one has determined constant matrices $A^{(1)}$, $A^{(2)}$, and $A^{(3)}$ in the approximations $VA^{(1)}$, $VA^{(2)}$, and $VA^{(3)}$ for the unknowns $U^{(1)}$, $U^{(2)}$, and $U^{(3)}$ in the problems

$$LU^{(1)} = P^{(1)} \,,$$
$$LU^{(2)} = P^{(2)} \,,$$
and
$$LU^{(3)} = P^{(3)} \,,$$

one can determine a constant A in an approximation VA for the unknown U in the problem

$$LU = P$$

by the relation

$$A = c^{(1)}A^{(1)} + c^{(2)}A^{(2)} + c^{(3)}A^{(3)} \quad ,$$

provided the relation

$$P = c^{(1)}P^{(1)} + c^{(2)}P^{(2)} + c^{(3)}P^{(3)}$$

holds and the constants $c^{(1)}$, $c^{(2)}$, and $c^{(3)}$ are known. The equations,

$$\begin{aligned} LU &= P \\ &= c^{(1)}P^{(1)} + c^{(2)}P^{(2)} + c^{(3)}P^{(3)} \\ &= c^{(1)}LU^{(1)} + c^{(2)}LU^{(2)} + c^{(3)}LU^{(3)} \\ &= L\left(c^{(1)}U^{(1)} + c^{(2)}U^{(2)} + c^{(3)}U^{(3)}\right) \quad , \end{aligned}$$

show that the unknown U is the same as the expression $c^{(1)}U^{(1)} + c^{(2)}U^{(2)} + c^{(3)}U^{(3)}$. If the approximations are substituted for $U^{(1)}$, $U^{(2)}$, and $U^{(3)}$, there results an approximation $V(c^{(1)}A^{(1)} + c^{(2)}A^{(2)} + c^{(3)}A^{(3)})$, say VA, for U.

The constant A determined by a superposition of constants is the same as the one which would be obtained by the method of functionals, provided the same functional F is used in every one of the approximations. One easily verifies the statement

$$\begin{aligned} A &= (FLV)^{-1}FP \\ &= c^{(1)}(FLV)^{-1}FP^{(1)} + c^{(2)}(FLV)^{-1}FP^{(2)} + c^{(3)}(FLV)^{-1}FP^{(3)} \\ &= c^{(1)}A^{(1)} + c^{(2)}A^{(2)} + c^{(3)}A^{(3)} \end{aligned}$$

under the assumption of the existence of $(FLV)^{-1}$. In the alternative situation the matrix A is exactly the same function as $c^{(1)}A^{(1)} + c^{(2)}A^{(2)} + c^{(3)}A^{(3)}$ of the same arbitrary elements, provided the transition from (FLV) to (GMW) is made in the same way in all four approximations.

When P is an infinite expansion in terms of $P^{(1)}$, $P^{(2)}$, $P^{(3)}$, ... , the usual questions of convergence arise. If $P^{(1)}$, $P^{(2)}$, $P^{(3)}$, ... are the functions of a complete set for expansions, the value of the idea of superposition of approximations is augmented.

4. Inclusion of Arbitrary Parameters.

If V or F includes one or more parameters, the matrix A is solved for in terms of them. It is then possible to vary the parameters and minimize the expression chosen to measure the error. Obviously the expression selected for the measurement of error has much to do with conclusions concerning magnitude of the error.

Variations can be applied to V or F or to both. Let κ be an arbitrary number and $\bar{V}$ be a matrix similar to V. If V is replaced by the varied matrix $V + \kappa\bar{V}$, one gets the equations

$$(FLV + \kappa FL\bar{V})A = FP$$

for the determination of the matrix A, each element of which is a rational fractional function in κ. Although this result appears rather complicated, it indicates the way in which variations in V can be studied. Let the matrix $F + \nu\bar{F}$

be a varied value of F, ν an arbitrary number, and $\bar{F}$ a functional similar to F. In the absence of a variation in V one gets the equations

$$(FLV + \nu \bar{F}LV)A = FP + \nu \bar{F}P$$

for the determination of the matrix A, each element of which is now a rational fractional function in ν. Variations in F can be studied with this beginning. If variations of the above nature occur simultaneously, the equations

$$(FLV + \kappa FL\bar{V} + \nu \bar{F}LV + \kappa\nu \bar{F}L\bar{V})A = FP + \nu \bar{F}P$$

are obtained for the determination of the resulting matrix A. The expressions are apparently too complicated for practical use in the choice of V and F, unless the matrices F, L, V, and P are such that much simplification is obtained. There are possibilities in the examination of expressions of measure of error under variations of the nature described.

There is the possibility of assuming that A included as elements variables in certain ways such that one can obtain the results of the application of FL upon VA and be able to arrive at a solution for A of the type assumed. Investigation of this is beyond the scope of this paper.

5. <u>Method of Functional Equations</u>.

There is one possibility for the functional F of which the reader must be reminded. The functional

$$F(\quad) = [10]_p(\quad) \ ,$$

in which one evaluates an expression at a point p of its region of definition, is very practical in applications. There are no integrations or differentiations to become troublesome in obtaining the numerical values of the elements of the matrices FLV and FP. An increase in the number of independent variables of a problem does not necessitate a search for another method of solution. Complex conditions over complicated regions can be approximated, although the solution of such problems by ordinary methods is generally hopeless.

There is a rather artificial kind of expansion $U \cong VA$ having, however, more than formal significance. It is an expansion appropriate for the expression of a method in which the values of the unknown U at a finite number of points p are approximated as an unknown constant matrix $\mathcal{U}$. Although one has the approximation $U \cong \mathcal{U}$, one can write it $U \cong I\mathcal{U}$. Because the matrix I is the unit matrix, its elements are known functions and one has a special case of V. Because the elements of $\mathcal{U}$ are constants to be determined, one has a special case of A.

Since L is often an operator over a continuum, the operation $L(I\mathcal{U})$ must be replaced by $\mathcal{L}(I\mathcal{U})$ in which $\mathcal{L}$ is a corresponding operator over the points p. After the operation $\mathcal{L}$ has been performed, one has $\mathcal{L}I\mathcal{U} \cong P$.

The unknown $\mathcal{U}$ is obtained after the application of the functional

$$F(\ ) = [1()]_p(\ )$$

from the equations

$$F\mathcal{L}I\mathcal{U} = FP\,.$$

One now has the relation of the method of functionals to methods of difference equations, to the method (9) of Fredholm (10) for the solution of integral equations, and to an extension the author has initiated and hopes to study in the future of the method of Fredholm to certain linear operational equations.

6. Replacement of Functional by a Linear Operator.

If F is selected from the field of linear operators, the elements of FLV and FP are not restricted to the field of constants and solution for A can not always be carried out. However, one can sometimes get a result as an example illustrates.

Consider the problem $(L-\lambda\bar{L})U=P$ for which L^{-1} is known. The Neumann series solution (8) by the method of successive approximation is

$$U = L^{-1}P + \lambda L^{-1}\bar{L}L^{-1}P + \lambda^2 L^{-1}\bar{L}L^{-1}\bar{L}L^{-1}P + \cdots$$
$$\cdots + \lambda^n (L^{-1}\bar{L})^n L^{-1}P + \cdots \quad .$$

It is necessary that $\lim\limits_{n\to\infty} \lambda^n (L^{-1}\bar{L})^n U$ vanish and that the series be convergent. In order to get this series by the method of functionals one takes

$$U \cong L^{-1}PA_0 + L^{-1}\bar{L}L^{-1}PA_1 + (L^{-1}\bar{L})^2 L^{-1}PA_2 + \cdots$$
$$\cdots + (L^{-1}\bar{L})^n L^{-1}PA_n \quad ,$$

$$F_i(\ ) = (L^{-1}\bar{L})^i(\ ) \qquad (i=0,1,2,\ldots,n) \quad ,$$

and neglects terms involving higher powers of $(L^{-1}\bar{L})$ than n. One obtains $A_i = \lambda^i$.

The solution (9) by successive approximation of the integral equation of Volterra of the second kind with a variable limit is a specialization of the procedure outlined in the preceding paragraph.

IV. DISCUSSION

Any method in which linear equations are used to determine unknowns appearing linearly in the solution of a linear problem is included in the method of functionals. The method of this paper generalizes and unites these methods. While some of the special cases of the linear operational equation $LU=P$ cannot be solved by other known methods, one is not only able to treat any special case but also is permitted much flexibility in the process. One is free to choose the functions of expansion for the unknown and may select the functional matrix in convenient ways.

One has the same flexibility in the solution for characteristic values in the problem $LU=P$. A solution can always be obtained. One is not halted by limitations of the method or by complexities introduced by a less flexible process.

Ways are given for estimating and measuring the error involved in the approximation by the method of functionals. When the functions of expansion are such that an exact answer can be obtained, it is found by the method of functionals. Ways of estimating the error involved are given in terms of expressions for the error permitted by the functions of expansion.

The specialization of $LU=P$ to examples promises a practical means of solution for many of them.

V. CONCLUSIONS

1. An approximate solution of a linear operational equation can be carried out by the use of functionals, although each step is made quite general in an effort to make the method as flexible as possible.

2. An estimation of the error involved in the approximate solution obtained by the method of functionals can be made in several ways.

3. If the functions of expansion are sufficient for a representation of the unknown, the approximation by the method of functionals is an exact solution.

4. Ways of estimating the error involved in the approximation by the method of functionals can be given in terms of expressions for the error permitted by the functions of expansion.

5. Although one can get numerical values for the characteristic values in a problem, it is difficult, at least without specializing the problem, to obtain an estimation of the error.

6. The relation of the method of functionals to many of

the methods for the approximate solution of special cases of the linear operational equation can be shown.

7. Although a problem may be such that no solutions of part of the conditions are known, one may still proceed toward an approximate solution by the method of functionals. Briefly, one may proceed, although part of the answer to the problem is not evident or accessible by existing methods.

8. Extensions of the method can be made to include biorthogonalization, an orthogonalization process, superposition of approximations, appearance of arbitrary parameters in the approximation, method of functional equations, and the method of successive approximation.

VI. SUMMARY

1. Several non-homogeneous, homogeneous, and characteristic value problems are given as examples to indicate the generality of a certain linear operational equation.

2. A formal approximation method is sketched for the solution of a problem $E(U;K)=0$ for comparison with the solution of the linear problem by the method of functionals.

3. An approximate solution of the linear operational equation $LU=P$ is carried out.

4. Ways of obtaining an estimation of the error involved in the approximate solution are given.

5. If the functions of expansion are sufficient for the representation of an exact answer of the problem, it is shown that an exact answer is obtained by the method of functionals.

6. Ways of estimating the error involved in the approximation by the method of functionals are given in terms of expressions for the error permitted by the functions of expansion.

7. Although one can get numerical values for the characteristic values in a problem, little material is given for the

estimation of the degree of approximation.

8. The relation of the method of functionals to methods of Ritz, Boussinesq (least squares), Krawtchouk, Schrödinger, Epstein, Trefftz, Taylor, and others is brought out.

9. Biorthogonalization is shown to involve the application of a non-diagonal type of functional. Advantages in computation of an analogue of orthogonalization of functions are shown.

10. If a linear operational equation is a linear combination of several problems, the approximate solution is shown to be the same linear combination of the approximate solutions of the several problems.

11. It is indicated that arbitrary parameters can be made to appear in the approximation. A means of giving the functions of expansion and the functionals variations is initiated.

12. It is shown how methods of difference equations, of Fredholm's solution of integral equations, and of an extension of Fredholm's method can be included formally in the method of functionals.

13. If the functional is replaced by a linear operator, one has agreement with the method of successive approxima-

tions. It is shown how the Neumann series solutions of problems can be obtained.

VII. LITERATURE CITED

(1) Ritz, W. Über eine neue Methode zur Lösung gewisser Variationsprobleme der mathematischen Physik. Journal für die reine und angewandte Mathematik, 135, 1909. pp. 1-61.

(2) Boussinesq, J. Théorie Analytique de la Chaleur. 1901, Volume 1. p. 316.

(3) Kravchuk (Krawtschouk), M. P. Sur la résolution des équations linéaires différentielles et intégrales par la méthode de moments. Kiïv, [Vidavnitstvo Ukraïns'koï akademiï nauk]. 1932-36.

In Ukrainian, with French resumé at end of each part.

(4) Pauling, L. and Wilson, E. B., Jr. Introduction to Quantum Mechanics. 1935, First Edition. pp. 151-93.

(5) Schrödinger, E. Über das Verhältnis der Heisenberg-Born-Jordanschen Quantenmechanik zu der meinen. Annalen der Physik, series 4, 79, Feb.-May 1926. pp. 734-56.

(6) Epstein, P. S. The Stark Effect from the Point of View of Schroedinger's Quantum Theory. Physical Review, series 2, 28, July-Dec. 1926. pp. 695-710.

(7) Trefftz, E. Konvergenz und Fehlerschätzung beim Ritzschen Verfahren. Mathematische Annalen, 100, Aug.-Dec. 1928. pp. 503-21.

(8) Courant, R. and Hilbert, D. Methoden der Mathematischen Physik. 1931, Second Revised Edition, Volume 1. pp. 41-2, 119-20.

(9) Goursat, E. Cours d'Analyse Mathématique. 1915, Second Edition, Volume 3. pp. 323-438.

(10) Fredholm, I. Sur une Classe d'Équations Fonctionnelles. Acta Mathematica, 27, 1903. pp. 365-90.

VIII. ACKNOWLEDGMENT

I wish to express my sincere gratitude to

Dr. J. V. Atanasoff

for his suggestion of the problem and for his guidance and assistance in its treatment.